A Remark On A Model Criticism Technique

by

Seymour Geisser

University of Minnesota

Technical Report No. 409

October 1982

A Remark on a Model Criticism Technique

by
Seymour Geisser*
University of Minnesota

Box (1980) suggests a method for criticizing an entertained model consisting of data y, parameter set θ and assumptions A, which is structured via the relationship,

$$p(y, \theta|A) = p(y|\theta,A)\ p(\theta|A).$$

Here $p(y|\theta,A)$ is the joint probability function of the observations given θ and $p(\theta|A)$ is the prior probability function of θ. He notes that prior to the availability of the data one can compute

$$p(y|A) = \int p(y|\theta,A)\ p(\theta|A)\ d\theta$$

which he denotes as the predictive (marginal) density of the data. This, he claims, enables one to assess the credibility of the model for any observed set of data y_d by referring to $p(y_d|A)$ or to the density $p(g(y_d)|A)$ of some predictive checking function $g(y_d)$.

Basically he defines a test with significance level

$$\alpha = \Pr\{p(y|A) < p(y_d|A)\}$$

to allow criticism of the model. He illustrates the concept by presenting several useful examples. In this note we shall present two examples which when taken in tandem throw some doubt on an uncritical use of this procedure.

Assume an i.i.d. sequence of Bernoulli trials with probability θ of success Suppose in this instance the prior denisty for θ is actually assumed to be uniform as in Bayes' original model. Then for a fixed number n of trials where y successes are observed, the predictive probability function of y

*Supported in part by NIH grant GM25271

is easily calculated to be

$$\Pr(y|A) = \frac{1}{n+1} \qquad y = 0, 1, \ldots, n. \tag{1}$$

i.e. uniform for all admissible values of y. Hence no significance test of the type advocated by Box,

$$\Pr\{p(y|A) < p(y_d|A)\} = \alpha \tag{2}$$

is available. Are we to conclude that Bayes' original model cannot be flawed? Or is it just that predictive model criticism fails here?

Suppose now we had used negative binomial sampling so that we terminated the experiment as soon as y successes were attained and consequently observed n trials. Here the predictive probability function of the number of trials is

$$\Pr(N=n|A) = \frac{y}{n(n+1)} \qquad n = y, y+1, \ldots . \tag{3}$$

The fact that the probability function is monotonically decreasing in n indicates that the Box procedure is workable i.e. if the observed $N = n_0$ is large enough relative to y, the model may be called into question. In fact,

$$\Pr[N \geq n_0] = \sum_{n=n_0}^{\infty} \frac{y}{n(n+1)} = \frac{y}{n_0} = \alpha, \tag{4}$$

where $\alpha = \hat{\theta}$, the MLE of θ. Are we then to conclude that predictive model criticism here succeeds only for small $\hat{\theta}$? Sampling until a fixed number of failures is attained results in criticism increasing with $\hat{\theta}$. In either case what aspect of the model is called into question other than Bayes' uniform prior? And, what is the meaning of calling this into question? Box has made an elegant suggestion for the problem of model criticism, but it should, like most statistical techniques, be used with caution and carefully interpreted within the context of its application.

REFERENCES

Box, G.E.P. (1980). Sampling and Bayes' inference in scientific modelling and robustness. J.R. Statist. Soc. A, 143, 383-430.